CS168: The Modern Algorithmic Toolbox
#12: Spectral Graph Theory, Part 2

Tim Roughgarden & Gregory Valiant*

May 5, 2021

In last lecture, we introduced the notion of the Laplacian matrix, L, associated to a graph. We defined it as $L = D - A$, where D is the diagonal matrix with element $D_{i,i}$ defined as the degree of the ith node, and A as the adjacency matrix of the graph. The key insight, which provided an intuitive understanding of the eigenvectors/eigenvalues of the Laplacian, was the calculation characterizing the associated quadratic form $v^T Lv$ as corresponding to assigning value $v(i)$ to the ith node in the graph, and then summing the squares of the differences in entries, across all edges of the graph:

$$v^T LV = \sum_{(i,j) \in edges} (v_i - v_j)^2.$$

This characterization led to our understanding that the eigenvectors of L with small eigenvalue tend to give similar values to neighboring nodes, and the eigenvectors with largest eigenvalues tend to give different values to neighboring nodes. These observations motivated using spectral embeddings that use the small eigenvectors for applications like 1) visualization, or 2) "spectral clustering". In contrast, embedding onto the largest eigenvectors is useful for applications like k-coloring, where neighboring vertices should be assigned to different sets/colors.

1 Conductance, isoperimeter, and the second eigenvalue

Last lecture, we leveraged the characterization of the quadratic form $v^T Lv$ to prove that the multiplicity of the zero eigenvalue is the number of connected components. Based on this, it seems intuitively clear that if λ_2 is extremely small, then the graph might be "close" to having two connected components, in the sense that there is a way of partitioning the nodes of the graph into two sets, with very few edges crossing from one set to the other. We now formalize this connection between λ_2, and the quality of the best such partition.

There are several natural metrics for quantifying the quality of a graph partition. We will state two such metrics. First, it will be helpful to define the *boundary* of a partition:

Definition 1.1 Given a graph $G = (V, E)$, and a set $S \subset V$, the *boundary* of S, denoted $\delta(S)$ is defined to be the set of edges of G with exactly one endpoint in set S.

One natural characterization of the quality of a partition $(S, V \backslash S)$ is the *isoperimetric ratio* of the set, defined to be the ratio of the size of the boundary of S to the minimum of $|S|$ and $|V \backslash S|$:

Definition 1.2 The *isoperimetric ratio* of a set S, denoted $\theta(S)$, is defined as

$$\theta(S) = \frac{|\delta(S)|}{\min(|S|, |V \backslash S|)}.$$

The *isoperimetric number* of a graph G is defined as $\theta_G = \min_{S \subset V} \theta(S)$.

A related notion of the quality of a partition, is the *conductance*, which is the ratio of the size of the boundary $\delta(S)$ to the minimum of the number of edges involved in S and the number involved in $V \backslash S$ (where an edge is double-counted if both its endpoints lie within the set). Formally, this is the following:

Definition 1.3 The *conductance* of a partition of a graph into two sets, $S, V \backslash S$, is defined as

$$cond(S) = \frac{|\delta(S)|}{\min(\sum_{i \in S} degree(i), \sum_{i \notin S} degree(i))}.$$

1.1 Connections with λ_2

There are connections between the second eigenvalue of a graph's Laplacian and both the isoperimetric number as well as the minimum conductance of the graph. The following surprisingly easily proved theorem makes the first connection rigorous:

Theorem 1.4 *Given any graph $G = (V, E)$ and any set $S \subset V$, $\theta(S) \geq \lambda_2(1 - \frac{|S|}{|V|})$. In other words, the isoperimetric number of the graph satisfies:*

$$\theta_G \geq \lambda_2(1 - \frac{|S|}{|V|}).$$

Proof: Given a set S, consider the associated vectors v_S defined as follows:

$$v_S(i) = \begin{cases} 1 - \frac{|S|}{|V|} & \text{if } i \in S \\ -\frac{|S|}{|V|} & \text{if } i \notin S \end{cases}$$

We now compute $\frac{v_S^t L v_S}{v_S^t v}$. First, we consider the numerator:

$$v_S^t L v_s = \sum_{i<j:(i,j) \in E} (v_S(i) - v_S(j))^2 = |\delta(S)|,$$

since the only terms in this sum that contribute a non-zero term correspond to edges with exactly one endpoint in set S, namely the boundary edges. We now calculate $v_S^t v_S = |S|(1-\frac{|S|}{|V|})^2 + (|V|-|S|)(\frac{|S|}{|V|})^2 = |S|(1-\frac{|S|}{|V|})$.

Hence we have shown that $\frac{v_S^t L v_S}{v_S^t v_S} = \frac{|\delta(S)|}{|S|(1-\frac{|S|}{|V|})}$. On the other hand, we also know that $\sum_i v_S(i) = 0$, hence $\langle v_S, (1,1\ldots,1)\rangle = 0$. Recall that

$$\lambda_2 = \min_{v:\langle v,(1,\ldots,1)\rangle=0} \frac{v^t L v}{v^t v} \leq \frac{v_S^t L v_S}{v_S^t v_S} = \frac{|\delta(S)|}{|S|(1-\frac{|S|}{|V|})}.$$

Multiplying both sides of this inequality by $(1-\frac{|S|}{|V|})$ yields the theorem. ■

What does Theorem 1.4 actually mean? It says that if λ_2 is large (say, some constant bounded away from zero), then *all* small sets $|S|$ have *lots* of outgoing edges—linear in $|S|$. For example, if $\lambda_2 \geq 1/2$, then for all sets S with $|S| \leq |V|/4$, we have that $\theta(S) \geq \frac{1}{2}(1-1/4) = 3/8$, which implies that $|\delta(S)| \geq \frac{3}{8}|S|$.

There is a partial converse to Theorem 1.4, known as *Cheeger's Inequality*, which argues that if λ_2 is small, then there exists at least one set S, such that the conductance of the set S is also small. Cheeger's inequality is usually formulated in terms of the eigenvalues of the *normalized Laplacian* matrix,defined by normalizing entry $L(i,j)$ by $1/\sqrt{deg(i)\cdot deg(j)}$. Rather than formally defining the normalized Laplacian, we will simply state the theorem for regular graphs (graphs where all nodes have the same degree):

Theorem 1.5 (Cheeger's Inequality) *If λ_2 is the second smallest eigenvalue of the Laplacian of a d-regular graph $G = (V,E)$, then there exists a set $S \subset V$ such that*

$$\frac{\lambda_2}{2d} \leq cond(S) \leq \frac{\sqrt{2\lambda_2}}{\sqrt{d}}.$$

1.2 Beyond λ_2?

Is there an analog of Theorem 1.4 and Cheeger's inequality that applies to λ_3, or higher eigenvalues? If $\lambda_k = 0$, we know the graph has k connected components, and hence it is tempting to conclude that if λ_k is nonzero, but small, the graph should have a partition into k pieces with few edges crossing between partitions. In some sense, this has been well-known empirically, though it was only in the past decade that rigorous analogs of "higher order" Cheeger's inequality were established [1], and this is an area of active research.

2 Random walks and diffusion on graphs

Beyond their use for graph partitioning and visualization, the eigenvalues and eigenvectors of a graph arise naturally when considering random processes on graph. Below, we describe one concrete model of diffusion over a graph, and give a high-level discussion of how the

eigenvalues and eigenvectors connect to this process. Next week, when we discuss Markov Chains, we will expand on some of these connections.

Consider the following basic model of diffusion, which models a variety of biological phenomena, as well as how beliefs, political views, cultural norms, etc. spread in a social network: Let A denote the adjacency matrix of a graph, $G = (V, E)$ with $|V| = n$, and assume that at time $t = 1$, the vector v_1 has entries corresponding to the initial views of each of the n nodes. (For example, let $v_1(i) \in (0, 1)$ represent how strongly person i believes that people should wear masks in public spaces.) Consider the following dynamics: at each time $t > 1$, the views of each node are replaced by the average of the views that their neighbors had at time $t - 1$. Formulating this update as a matrix product, we have:

$$v_{t+1} = D^{-1}Av_t,$$

where D^{-1} denotes the diagonal matrix whose ith diagonal entry is $1/deg(i)$.

These dynamics do not necessarily converge in the limit as t gets large: for a network with just two people with an edge between them, they will "swap" their views at each timestep. Nevertheless, *if* the dynamics do eventually converge to a vector, v, we know that vector must satisfy $v = D^{-1}Av$, and hence v will be an eigenvector of $D^{-1}A$ with eigenvalue 1!!

In the case where this convergence happens, how quickly will it be expected to occur? Namely, how large a value of t will be necessary for $v_t \approx v$? Recall from our analysis of the power iteration algorithm from Lecture 8, that the diffusion update defined above is simply performing the power iteration algorithm that computes the top eigenvector of $D^{-1}A$—the only difference is that in the power iteration algorithm, we choose a uniformly random unit vector as the initial vector, whereas in the above dynamics, the initial vector v_1 corresponds to the initial "views" of the corresponding nodes. In our analysis of the power iteration algorithm, we saw that the ratio of the largest to second largest eigenvalues determined how quickly we would expect the dynamics to converge to the top eigenvector. (If the eigenvalues are very close, then we will need to run for more iterations.)

At a high level, if the dynamics are run on a well-connected graph, then 1) they will converge quickly (if they do actually converge) 2) the ratio of the largest to second largest eigenvalue of $D^{-1}A$ will be large, and 3) the second smallest eigenvalue of the Laplacian will be large. Conversely, if the graph is not well-connected (e.g. a graph consisting of two clusters with only a single edge crossing between them), then the dynamics will converge slowly, the second largest eigenvalue of $D^{-1}A$ will be similar to the largest, and the second-smallest eigenvalue of the graph Laplacian will be close to zero. These high-level principles will be useful to keep in mind next week, as we discuss Markov Chains and Markov Chain Monte Carlo (MCMC).

References

[1] James R Lee, Shayan Oveis Gharan, and Luca Trevisan. Multiway spectral partitioning and higher-order Cheeger inequalities. *Journal of the ACM (JACM)*, 61(6):37, 2014.